Workgroup: MOPS
Internet-Draft: draft-giuliano-treedn-02
Published: 2 December 2022
Intended Status: Informational
Expires: 5 June 2023
Authors: L. Giuliano C. Lenart R. Adam
Juniper Networks Verizon GEANT

TreeDN- Tree-based CDNs for Live Streaming to Mass Audiences

Abstract

As Internet audience sizes for high-interest live events reach unprecedented levels and bitrates climb to support 4K/8K/AR, live streaming can place a unique type of stress upon network resources. TreeDN is a tree-based CDN architecture designed to address the distinctive scaling challenges of live streaming to mass audiences. TreeDN enables operators to offer Replication-as-a-Service (RaaS) at a fraction the cost of traditional, unicast-based CDNs- in some cases, at no additional cost to the infrastructure. In addition to efficiently utilizing network resources to deliver existing multi-destination traffic, this architecture also enables new types of content and use cases that previously weren't possible or economically viable using traditional CDN approaches. Finally, TreeDN is a decentralized architecture and a democratizing technology for content distribution.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 5 June 2023.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction

Live streaming to mass audiences can impose unique demands on network resources. For example, live sporting events broadcast over the Internet to end users has much lower tolerance for long playout buffers than typical on-demand video streaming. Viewers of live sporting events have long been conditioned by broadcast television to expect to see the content in real time, with only very short buffers for broadcast delays to prevent profanity and other objectionable content from making on the air (the "seven-second delay"). With micro-betting, even this 5-10 second delay can be too long. By comparison, when watching on-demand movies, an extra one- or two-minute playout buffer tends to be perfectly acceptable for viewers. If playout buffers for live sports are that long, viewers run the risk of being alerted to the game winning score from text messages from friends or cheers from the bar across the street, minutes before they view it themselves.

Another unique characteristic of live streaming is join rate. While on-demand video streaming can consume massive amounts of network resources, the viewing rates tend to be smooth and predictable. Service Providers observe gradual levels of traffic increases over the evening hours corresponding to prime-time viewing habits. By comparison, viewing rates of live video streams can more closely resemble step functions with much less predictability as mass audiences of viewers tune in to watch the game at the same time.

Previous efforts at more efficient network replication of multi-destination traffic have experienced mixed success in terms of adoption. IP multicast is widely deployed on financial networks, video distribution networks, L3VPN networks and certain enterprises. But most of these deployments are restricted to "walled-garden" networks. Multicast over the global Internet has failed to gain traction, as only a very small portion of the Internet is multicast-enabled at this time.

TreeDN is the result of the evolution of network-based replication mechanisms based on lessons learned from what has and has not worked well in the past. TreeDN addresses the fundamental issues of what has hindered multicast from adoption on the global Internet and enables service providers the opportunity to deliver new Replication-as-a-Service (RaaS) offerings to content providers, while more efficiently utilizing network resources, and thus, improving the experience of end users. Further, by more efficiently supporting multi-destination traffic, TreeDN is an architecture that can enable new types of content, such as Augmented Reality (AR) live streaming to mass audiences, that previously weren't possible or economically viable on the Internet due to the inefficiencies of unicast.

1.1. Requirements Language

The key words "**MUST**", "**MUST NOT**", "**REQUIRED**", "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**RECOMMENDED**", "**NOT RECOMMENDED**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Applicability

While the primary use case mentioned throughout this document is live streaming of multimedia content (audio, video, AR, real-time telemetry data), the TreeDN architecture is ideal for any content that needs to be replicated and delivered to multiple destinations. For example, large software file updates (eg, OS upgrades) that need to be delivered to many end users in a very short window of time can cause significant strain on network resources. Using TreeDN, this use case be handled much more efficiently by the network.

3. Problem Statement

The following issues have been the primary challenges for deployment of IP multicast over the global Internet:

* The "All or Nothing" Problem: IP multicast requires every layer 3 hop between source and receivers to be multicast-enabled. To achieve ubiquitous availability on the global Internet, this essentially means nearly every interface on every router and firewall on the Internet must support a multicast routing protocol like PIM-SM [RFC7761] or mLDP [RFC6388]. This requirement creates a bar to deployment that is practically impossible to overcome.

* The "It's Too Complex" Problem: operators have long complained that multicast routing protocols like PIM-SM are simply too complex, making it costly to design, configure, manage and troubleshoot IP multicast in the network.

* The "Chicken and Egg" Problem: there's not much multicast content because there's not much of a multicast-enabled audience, but there's not much of a multicast-enabled audience because there's not much multicast content.

TreeDN is the evolution of network-based replication based on lessons learned over decades and is designed to address the problems listed above.

4. TreeDN Architecture

TreeDN leverages advances in the availability and understanding of overlay and underlay networking. With network overlays, a service can be achieved and delivered to end users while recognizing and tolerating the practical realities of what is possible over a network as diverse as the global Internet. That is, the replication service is available to users and applications regardless of what protocols may exist in the underlying networks.

4.1. TreeDN Overlays

One overlay technology that TreeDN leverages is Automatic Multicast Tunneling (AMT) [RFC7450]. With AMT, users on unicast-only networks (AMT Gateways) can dynamically build tunnels to routers on the multicast-enabled part of the network (AMT Relays) and receive multicast streams. The AMT Gateway is a thin software client which typically sits on the receiving end host and initiates the tunnel at an AMT Relay, which is a tunnel server that typically sits at the border of the multicast network. AMT allows any end host on the Internet to receive multicast content regardless of whether their local provider supports multicast (aka, "off-net receivers"), which addresses the "All or Nothing" Problem. Links and devices that do not support multicast are simply tunneled over- they no longer present a barrier to the overall replication service for end users. Those networks that do deploy and support multicast, as well as the content providers that serve up multicast content, are able to enjoy the benefits of efficient replication and delivery. Further, these benefits can serve as incentives for operators who do not yet support multicast to enable it on their networks. Once the cost of carrying duplicated unicast tunnels is perceived by those operators to exceed the cost of deploying multicast, they are more likely to enable multicast on their networks. In this way, TreeDN effectively supports incremental deployment in a way that was not previously possible with traditional (non-overlay) multicast networking. Finally, AMT also addresses the "Chicken and Egg" Problem, as all end hosts on the global Internet that have access to an AMT Relay are potential audience members.

In addition to AMT, other overlay technologies like Locator/ID Separation Protocol (LISP) [RFC6830] can be utilized to deliver content from multicast-enabled networks to end hosts that are separated by portions of the network (at the last/middle/first mile) that do not support multicast.

4.2. TreeDN Native On-Net

Networks that support multicast provide the native on-net component of TreeDN. The primary requirement of the native on-net is to support Source-Specific Multicast (SSM) [RFC4607]. PIM-SSM, which is merely a subset of PIM-SM, is the multicast routing protocol typically used in SSM. However, any multicast routing protocol capable of supporting SSM can be used as a TreeDN native on-net, such as mLDP, GTM [RFC7716] and BGP-based Multicast [I-D.ietf-bess-bgp-multicast], or even BGP-MVPN [RFC6513] for those operators who carry the global routing table in a VRF. Likewise, any data plane technology that supports SSM, including BIER [RFC8279] and SR-P2MP [I-D.ietf-spring-sr-replication-segment] can be used.

The key benefit of SSM as the native on-net component of TreeDN is that it radically simplifies the control plane needed to support replication in the network. This benefit addresses the "It's Too Complex" Problem. Most of the complexity of multicast is eliminated in SSM, which reduces the cost of deploying and operating a multicast network. Further rationale for this SSM-only approach can be found in ASM Deprecation [RFC8815].

5. Replication-as-a-Service (RaaS)

Content providers have traditionally used CDNs to distribute content that needs to be delivered to large audiences, essentially outsourcing the task of replication to CDN providers. Most CDNs utilize unicast delivery, as multicast is not an option due to its lack of general availability on the global Internet. TreeDN is a CDN architecture that leverages tree-based replication to more efficiently utilize network resources to deliver simultaneous multi-destination traffic. By leveraging overlay networking to address the "All or Nothing" and "Chicken and Egg" Problems and SSM to address the "It's Too Complex" Problem, TreeDN avoids the practical issues that previously prevented multicast from being a viable option for CDN providers.

TreeDN has several advantages over traditional unicast-based CDN approaches. First, the TreeDN functionality can be delivered entirely by the existing network infrastructure. Specifically, for operators with routers that support AMT natively, multicast traffic can be delivered directly to end users without the need for specialized CDN devices, which typically are servers that need to be racked, powered and connected to revenue-generating ports on routers. In this way, SPs can offer new RaaS functionality to content providers at potentially zero additional cost in new equipment (modulo the additional bandwidth consumption).

Additionally, TreeDN is an open, standards-based architecture based on mature, widely implemented protocols. TreeDN also requires far less coordination between the content provider and the CDN operator. That is, there are no storage requirements for the data, nor group-key management issues since a TreeDN provider merely forwards packets. A TreeDN provider simply needs to have enough accounting data (eg, traffic data, number of AMT tunnels, etc) to properly bill customers for the service. By contrast, traditional unicast-based CDNs often incorporate proprietary, non-interoperable technologies and require significant coordination between the content provider and the CDN to handle such things as file storage, data protection and key-management.

6. Decentralization/Democratization of Content Sourcing

TreeDN is an inherently decentralized architecture. This reduces the cost for content sourcing, as any host connected to a multicast-enabled network, or on a source-capable overlay, can send out a single data stream that can be reached by an arbitrarily large audience. By effectively reducing to zero the marginal cost to the source of reaching each additional audience member, TreeDN democratizes content sourcing on the Internet.

7. Integration with Unicast

Since SSM inherently implies unidirectional traffic flows from one to many, mechanisms that rely on bidirectional communication between receivers and the content provider, such as bespoke advertising, telemetry data from receivers detailing end user experience, distribution of decryption keys, switching to higher/lower bandwidth streams, etc, are not well suited to SSM delivery. As such, separate unicast streams between receivers and content providers may be used for this type of "out-of-band" functions while SSM is used to deliver the actual content of interest. Further details on this hybrid unicast-multicast model for content delivery are beyond the scope of this document.

8. Security Consideration

TreeDN is essentially the synthesis of SSM plus overlay networking technologies like AMT. As such, the TreeDN architecture introduces no new security threats that aren't already documented in SSM and the overlay technologies that comprise it. Further, RFC 4609 and RFC 8815 describes the additional security benefits of using SSM instead of ASM.

9. IANA Considerations

This document has no IANA actions.

10. Acknowledgements

Many thanks to those who have contributed to building and operating the first TreeDN network on the Internet, including Pete Morasca, William Zhang, Lauren Delwiche, Natalie Landsberg, Wayne Brassem, Jake Holland, Andrew Gallo, Casey Russell, Janus Varmarken, Csaba Mate, Frederic Loui, Max Franke, Todor Moskov and Erik Herz. The writing of this document to describe the TreeDN architecture was inspired by a conversation with Dino Farinacci and Mike McBride. Thanks also to Jeff Haas and Vinod Kumar for their thoughtful reviews and suggestions.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

11.2. Informative References

[I-D.ietf-bess-bgp-multicast]
Zhang, Z. J., Giuliano, L., Patel, K., Wijnands, I., Mishra, M. P., and A. Gulko, "BGP Based Multicast", Work in Progress, Internet-Draft, draft-ietf-bess-bgp-multicast-04, 7 January 2022, <https://www.ietf.org/archive/id/draft-ietf-bess-bgp-multicast-04.txt>.

[I-D.ietf-spring-sr-replication-segment]
Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. J. Zhang, "SR Replication Segment for Multi-point Service Delivery", Work in Progress, Internet-Draft, draft-ietf-spring-sr-replication-segment-10, 20 October 2022, <https://www.ietf.org/archive/id/draft-ietf-spring-sr-replication-segment-10.txt>.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, DOI 10.17487/RFC4607, August 2006, <https://www.rfc-editor.org/info/rfc4607>.

[RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011, <https://www.rfc-editor.org/info/rfc6388>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC6830] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "The Locator/ID Separation Protocol (LISP)", RFC 6830, DOI 10.17487/RFC6830, January 2013, <https://www.rfc-editor.org/info/rfc6830>.

[RFC7450] Bumgardner, G., "Automatic Multicast Tunneling", RFC 7450, DOI 10.17487/RFC7450, February 2015, <https://www.rfc-editor.org/info/rfc7450>.

[RFC7716] Zhang, J., Giuliano, L., Rosen, E., Ed., Subramanian, K., and D. Pacella, "Global Table Multicast with BGP Multicast VPN (BGP-MVPN) Procedures", RFC 7716, DOI

10.17487/RFC7716, December 2015, <https://www.rfc-editor.org/info/rfc7716>.

[RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <https://www.rfc-editor.org/info/rfc7761>.

[RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <https://www.rfc-editor.org/info/rfc8279>.

[RFC8815] Abrahamsson, M., Chown, T., Giuliano, L., and T. Eckert, "Deprecating Any-Source Multicast (ASM) for Interdomain Multicast", BCP 229, RFC 8815, DOI 10.17487/RFC8815, August 2020, <https://www.rfc-editor.org/info/rfc8815>.

Authors' Addresses

Lenny Giuliano
Juniper Networks
2251 Corporate Park Drive
Herndon, 20171
United States of America

Email: lenny@juniper.net

Chris Lenart
Verizon
22001 Loudoun County Parkway
Ashburn, 95134
United States of America

Email: chris.lenart@verizon.com

Rich Adam
GEANT
City House
126-130 Hills Road
Cambridge
CB2 1PQ
United Kingdom

Email: richard.adam@geant.org